

Accuracy and Efficiency of an Automated System for Calculating APACHE II Scores in an Intensive Care Unit

Valerie J. Gooder R.N., M.S., Brad R. Farr M.D., Ph.D., Michael P. Young, M.D.
McKay-Dee Hospital Center, Ogden, Utah

We evaluated the reliability and efficiency of an automated system for calculating APACHE II scores. We imported an automated APACHE II scoring system developed at another institution. We scored a convenience sample of 50 consecutive intensive care unit (ICU) admissions using three methods: (1) the automated system (2) an expert scorer using a manual data abstraction method, and (3) the current manual scoring method used in the ICU. We analyzed interrater reliability among the three groups, and compared scoring time between the automated and the expert groups. Interrater reliability testing demonstrated a very high agreement between the computer and the expert scorers, ($r = 0.97$ $p < 0.01$) and a high agreement between the computer and the nursing staff ($r = 0.80$, $p < 0.01$). The mean time required to complete the data sheet manually and input data into a standalone computer manually was 4 minutes and 55 seconds, compared to total mean time of 33 seconds for the automated system. Automated APACHE II scoring yielded results more reliable than those of an expert scorer, as judged by a second researcher. The time required to score patients was reduced and reliability of scores was improved with use of an automated APACHE II scoring system.

INTRODUCTION

"The complexity of modern medicine exceeds the inherent limitations of the unaided human mind."^{1p1272} This observation is well illustrated by an evaluation of systems commonly used to measure severity of illness of patients requiring intensive care. Severity of illness scores are calculated based on worst values for multiple physiologic variables. During a patient's stay in an intensive care unit (ICU) thousands of data values are recorded hourly by the bedside monitors, laboratories, and patient-care providers. Much of these data are used to calculate severity of illness scores.

To be useful, severity scoring must be accurate.² Humans are error prone when performing complex data collection and analysis. Any system that requires error-free performance from humans inevitably fails.³ Allowing an ICU staff to employ scoring systems successfully requires simplifying the task of data collection and analysis.⁴ An automated system that collects data for severity of illness classification may

reduce human error and improve efficiency of data collection, and thus may exceed the reliability of an expert manually entering and analyzing data.

BACKGROUND

Severity of illness classification for outcome prediction in ICUs is essential in today's health-care environment. Severity of illness classifications provide methods for determining cohort groups in research,⁵ evaluating new therapies, and for identifying quality improvement markers. In addition, severity of illness classification assists in stratification of patients for efficient use of expensive resources.^{6,7} We predict that researchers and clinicians will routinely require outcome data and risk factor adjustments to be part of the future patient-care record.⁸

APACHE II is a severity of illness scoring system used widely to predict the risk of death for ICU patients. The score given is numerical and ranges from 0 to 71. The score is calculated from 12 weighted physiologic variables obtained from data during the first 24 hours that a patient is in the ICU. APACHE II scoring was validated in 1985 using 5815 ICU admissions in 13 tertiary care centers.⁹

APACHE II has advantages when compared to other severity of illness scoring systems. APACHE II introduced mortality prediction into the clinical mainstream. It is used for administration, planning, quality assurance and resource allocation.¹⁰ In addition, APACHE II was used to compare outcomes among hospitals in relation to organizational differences.¹¹ Furthermore, APACHE II is used as a principal tool to measure severity of illness in multiple clinical trials evaluating new ICU therapies.¹² The APACHE II classification system is thus well described and has been validated for multiple patient types.^{2,13,14,15} Finally, the APACHE II system is in the public domain and it is available at no cost.

At McKay-Dee Hospital Center, APACHE II is used to evaluate the severity of illness of the ICU population. APACHE II scores are used to match cases by severity of illness and by disease classification when evaluating interventions and protocols.

The limitations of the APACHE II scoring system also are well described. APACHE II scores accurately predict mortality for large populations of ICU patients, but are not reliable for use in individual

cases. APACHE II is subject to lead-time bias. Delays in admission and varying referral methods may result in inaccurate outcome prediction.¹⁶ An additional limitation of all severity of illness scoring systems is they require substantial use of human resources for data collection. Because of resource constraints, severity of illness scores are not obtained routinely in many hospitals.

The number of data points reviewed for completion of APACHE II scores is extensive. Scoring of one randomly selected ICU admission at our institution required analysis of more than 400 data points. The large number of data points required raises concerns about scoring accuracy. Human short-term memory is limited to five to nine chunks, where chunk is defined as a unit of related information that is familiar from repeated encounters.¹⁷ Individual data elements may not have associations and thus function as individual chunks in memory.¹⁸ Information is held in working memory for only 15 to 20 seconds.²⁰ A data collector must scan data to determine the high and low values required for APACHE II scoring. Information in working memory is easily lost due to interference from incoming information or environmental distraction.¹⁸ As a result, data abstraction for calculation of APACHE II scores is prone to frequent errors.

Computation of APACHE II scores also requires complex calculations.¹⁹ Each physiological parameter is placed in a category that provides the highest score on a scale from 0 to 4. For example, a mean arterial blood pressure of 150 mm Hg yields a 3 point score whereas a mean arterial blood pressure of 40 mm Hg is awarded 4 points. At our institution, a standalone desktop computer was available for calculating APACHE II scores after manual data collection forms were completed. However, the collector must select those data that yield the highest point data prior to entering the data into the computer. The data collector must determine whether the higher or lower value provides the highest score.

Computer-based patient records provide opportunities for automated severity of illness scoring systems.²⁰ A goal of automated systems is to provide data that are easy to retrieve and that can be used for a variety of clinical purposes.⁸ Automation of APACHE II scoring may provide several benefits, including more timely reporting of scores to permit real-time use of the APACHE II data.

McKay-Dee Hospital uses the HELP (Health Evaluation through Logical Processing) hospital information system. In the ICU, a data-collection subsystem automatically transfers physiologic data from bedside monitors to the HELP system. This subsystem uses a filtering algorithm to reduce recording of erroneous vital sign data, and to limit the number of

data stored in the patient record. The filtering algorithm samples vital sign data each minute for 15 minutes, discards the high and low values, and stores the median of the remaining values.²¹ The SunQuest Laboratory system provides laboratory results to the HELP System. HELP remains clinically active more often than 99.5% of the time.²² A computer terminal is located at the bedside in all ICU patient rooms.

Currently, APACHE II scores are completed by 48 ICU nurses at our facility. Ideally, the score is calculated immediately at the end of the first 24-hour period after a patient's admission to the ICU. Unfortunately, APACHE II scores often are not completed for 2 to 8 weeks due to time constraints, lack of knowledge by nursing staff, or oversight. A research assistant collected data for 90% of the APACHE II scores because the nursing staff did not complete them. The research assistant was frequently three months behind on scoring. Retrospective scoring required manually retrieving the patient charts from the medical record department. Delays likely increase error in scoring. An automated system can require a nurse to complete APACHE II scoring prior to the generating an end-of-shift report. This incentive to complete the scoring also provides a tracking mechanism to monitor scoring completion.

The APACHE II score is based on measurement of the 12 physiologic variables and Glasgow coma score, age and an evaluation of chronic health status. The physiologic variables measured include arterial pH, arterial PO₂, age, mean arterial blood pressure, heart rate, temperature, respiratory rate, serum sodium, serum potassium, serum creatinine, white blood cell count, hematocrit, serum bicarbonate, glasgow coma score, chronic health evaluation, and the presence of acute renal failure.

Age, mean arterial blood pressure, heart rate, temperature, respiratory rate, serum sodium, serum potassium, serum creatinine, white blood cell count, hematocrit, and serum bicarbonate are available on the HELP system. Arterial pH, arterial PO₂, estimated glasgow coma score, chronic health evaluation, and the presence of acute renal failure must be manually entered into the scoring system by the nurse.

Several calculations are required for an APACHE II score to be computed. A calculation is required to determine the alveolar-arterial gradient when the FiO₂ is greater than .50. The points assigned for an abnormal serum creatinine are doubled in the event of acute renal failure. Physiologic variables are assigned point values from 0 to 4 by the rater: 4 is the highest level of derangement and absent values are given a default score of zero. Our research hypothesis predicted that we could increase reliability and reduce resource consumption by automating APACHE II

scoring at our institution.

METHODS

We used a descriptive correlational design for the study. We retrospectively studied 50 consecutive admissions to the Critical Care Unit from 7/18/95 to 8/7/95. An automated APACHE II scoring system was developed at LDS Hospital in Salt Lake City, Utah.²¹ We exported the system to McKay-Dee Hospital Center, modifying it to meet the requirements of our local information system and to achieve strict compliance with the scoring guidelines as outlined by Knaus and associates.⁹

The Expert Score

An expert scorer (VG), familiar with the automated system, determined the date and time that each of 50 consecutive patients was admitted to the ICU. Using the medical record number, she located and manually reviewed the patient's history and physical report to determine an operative status and chronic health status. She examined the physician and nursing notes to locate the required Glasgow coma score. In addition, she determined the presence or absence of acute renal failure and determined the most abnormal arterial blood gas (ABG) results. All data were recorded on standardized paper data sheets used by the institution for APACHE II scoring.

Using the computer-based patient record, the scorer generated and manually reviewed a report from the computer-based patient record of all stored vital signs for the patient's first 24 hours in the ICU. This report included the systolic and diastolic blood pressures, respiratory rate, temperature and heart rate. In addition, the expert scorer manually reviewed computer-based printouts of all chemistry and hematocrit values obtained during the first 24 hours of ICU care. The physiologic values yielding the highest point value were selected and recorded. The completed datasheet results were input into a standalone program designed to store manually collected data and to calculate APACHE II scores.

The Automated Score on HELP

To calculate the automated score, the expert scorer accessed the APACHE II automated program on the HELP system. She selected the patient by entering the account number. The automated system provided the patient's admission date and displayed the 24 hour interval following ICU admission. The first screen required manual input of ABG results. The next screen required the nurse to input "Yes" or "No" to answer the question "Does the patient have Acute renal failure?". The third screen prompted entry of the estimated

Glasgow coma score. The final screen required identification of an operative status (elective postoperative, nonoperative, or emergent postoperative). A subsequent screen displayed a list of chronic health problems for evaluation by the scorer. The expert scorer entered data into the required areas of the program with date and time stamp for each entry. The automated system calculated the APACHE II score using data stored in the computer-based patient record. This automated APACHE II score was not displayed to the expert scorer.

The ICU Staff Score

During the study period, all ICU nurses continued to collect data for APACHE II scoring in the usual manner. On admission, APACHE II scoring sheets were posted outside patient rooms. The unit clerks manually entered the date and time of admission onto the sheets. Twenty-four hours after admission (or on patient discharge if prior to 24 hours after ICU admission), the patient's nurse was expected to complete the APACHE II data collection and to place the data sheet in a designated collection area. The data sheets were collected by a research assistant and were manually entered into the standalone program. Final scores were tabulated by the standalone system and stored.

Time to Score

The time required for the expert scorer to complete the scoring was recorded. The time for scoring by the expert was defined as the time required to analyze the data after collection and to enter the data into the standalone computer program. The expert scorer time did not include time used for gathering the necessary documents for data analysis. One researcher recorded the time required to score the patients on the new automated system. The time for automated scoring included data entry of the four data elements not currently available on the computer-based patient record. The scoring time using the automated system did not include time spent gathering necessary documents for data analyses. The expert scorer used identical data for several values for the automated and manual scoring methods including ABG results, estimated glasgow coma score, assessment of acute renal failure, and chronic health evaluation. These data points did not differ between the expert scorer and the automated system on HELP. Time for the expert scorer to complete APACHE II scoring both manually and using the automated system was recorded on 23 patients. Because of the inconsistency in scoring, we were unable to track scoring time for the ICU nurses. Time values do not reflect the time spent by the ICU nurses collecting data for analysis using the manual

method. In addition, the time spent by the research assistant to find charts and to score patients retrospectively was not included in the analysis.

Interrater Reliabilities

APACHE II scores for each patient obtained by the computer, the expert scorer and the ICU staff were compared. We calculated interrater reliabilities using Pearson's product moment correlations for the automated and manual scoring systems. When scores disagreed, a second researcher (BF) reviewed the data used for scoring for both systems and determined which was more accurate. We obtained Institutional Review Board approval for this study.

RESULTS

Patient ages ranged from 10 to 90 years old. Multiple diagnostic categories were included: cardiovascular surgery (N= 15), general surgical (N=13), cardiovascular medical (N=11), general medical (N= 7), and neurosurgical (N=4). We scored patients requiring less than 24 hours of ICU care (N=10) using physiologic parameters for the 24-hour period, regardless of patient location at the time of scoring.

APACHE II scores for the 50 patients ranged from 5 to 31. Twenty-eight (54%) of the 50 scores were identical for the expert using the manual method and the computer. Six (14%) varied by 1 point and 11(22%) were within 2 points. Thus, 45 (90%) of the 50 scores were within two points. Interrater reliabilities using Pearson's product moment correlations demonstrated a strong positive relationship between the computer and the expert scorers ($r= 0.97$, $p< 0.01$) and strong positive relationship between the computer and the ICU staff ($r= 0.80$, $p< 0.01$). For the 22 disagreements, the computer was correct 21 (95.4%) of the time according to a second reviewer (BF). In 20 cases the automated score gave a higher score which is reflected in the higher overall mean score (Table 1).

Table 1: Comparison of scores generated by three methods.

Scoring method	n	M	SD
Automated	50	14.40	6.24
Expert	50	13.76	5.94
ICU Staff	50	13.62	7.22

The average time required to complete the

manual data sheet and to input the data was 4 minutes 55 seconds. Using the automated system, the average time required for total computer data entry was 33 seconds, inferring a mean time savings of 4 minutes 20 seconds per patient.

DISCUSSION

Automated APACHE II scoring yielded results more reliable than an expert manual scorer as judged by a second researcher (BF). The computer is able to scan numerical data reliably and quickly. In addition, the need to choose the physiologic derangement yielding the highest point value adds complexity to the calculations. The computer quickly assigns the maximum value for each physiologic variable.

The time required to score patients was also reduced by the use of an automated APACHE II scoring system. Future reductions in scoring time will be possible with automated ABG value entry.

Differences among scores on several patients can be explained by differences in methods of analysis of mean arterial pressure. To determine mean arterial pressure using the manual data sheet, the expert scorer uses systolic and diastolic pressures to calculate mean arterial pressure. The automated system, however, records mean arterial pressure from waveform analysis on the HP monitor. This feature results in potentially different numeric values being assigned to these parameters.

Our automated system demonstrated clear advantages over the manual system for calculation of APACHE II scores. Poor compliance by care givers is undesirable. The salary of the APACHE II research assistant required for retrospective data collection is an added expense eliminated by use of an automated system. Compliance with an automated system can be ensured if nurses are required to complete APACHE II scores in order to obtain the mandatory end-of-shift report. The expert scorer was accurate. However, we assume that the data are less reliable when manually collected by a large number of care givers.

One limitation of the system is the dependence of our automated scoring system on the HELP system and on the nearly completely computer-based patient record at McKay-Dee Hospital Center. The hierarchical database provides rapid retrieval of data for examination. Hospitals without a sophisticated computer-based patient record may find it difficult to employ an automated scoring system.

The automated APACHE II scoring system was implemented in the Intensive Care Unit at McKay-Dee Hospital in September 1995. The automated system continues calculating APACHE II scores on all admissions.

CONCLUSION

Our automated system for APACHE II scoring reduces staff time required for scoring and provides more timely scores, compared to manual scoring. In addition, our automated APACHE II system provides reliability of scores exceeding those of an expert scorer.

Acknowledgments

The authors wish to thank Thomas Oniki M.S. for development of the automated APACHE II system at LDSH, Salt Lake City, UT, and Dennis Bowden B.S. for modification and implementation of the system at McKay-Dee Hospital Center.

References

1. Eddy DM. Practical policies: Where do they come from? *JAMA* 1990;263:1265-75.
2. Ludwigs U, Hulting J. Acute physiology and chronic health evaluation scoring system in acute myocardial infarction: A prospective validation study. *Crit Care Med* 1995;23:854-59.
3. Leape LL. Error in medicine. *JAMA* 1994;272:1851-57.
4. Zimmerman JE, Knaus WE, Seneff M. *Intensive Care Rounds: Outcome prediction in Intensive Care*. Oxfordshire, England: The Medicine Group Ltd, 1993.
5. Knaus WA, Zimmerman JE, Wagner DP, Lawrence DE. APACHE: Acute physiology and chronic health evaluation: A physiologically based classification system. *Crit Care Med* 1981;9:591-97.
6. Barriere SL, Lowry SF. An overview of mortality risk prediction in sepsis. *Crit Care Med* 1995;23:376-93.
7. Kalb PE, Miller DH. Utilization strategies for intensive care units. *JAMA* 1989;261:2389-95.
8. Committee on Improving the Patient Record, Institute of Medicine. *The Computer-Based Patient Record: An Essential Technology for Health Care*. Washington: Institute of Medicine, 1991.
9. Knaus WA, Draper EA, Zimmerman JE. APACHE II: A severity of disease classification system. *Crit Care Med* 1985;13:818-29.
10. Carlson RW, Geheb MA editors. *Principles and Practice of Medical Intensive Care*. Philadelphia: W.B. Saunders Company, 1993.
11. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med* 1986;104:410-18.
12. Ziegler EJ, Fischer CJ, Sprung CL, et al. Treatment of gram negative bacteremia and septic shock with HA-1A human monoclonal antibody against endotoxin. *N Engl J Med* 1991;324:429-35.
13. Berger MM, Marazzi A, Freeman J, Chioloro R. Evaluation of the consistency of acute physiology and chronic health evaluation (APACHE II) scoring in a surgical intensive care unit. *Crit Care Med* 1992;20:1681-87.
14. Turner JS, Mudalier YM, Chang RWS, Morgan CJ. Acute physiology and chronic health evaluation (APACHE II) scoring in a cardiothoracic intensive care unit. *Crit Care Med* 1991;19:1266-69.
15. Castella X, Artigas A, Bion J, Kari A. A comparison of severity of illness scoring systems for intensive care unit patients: Results of a multicenter, multinational study. *Crit Care Med* 1995; 23:1327-35.
16. Roberts AH, Zimmerman J E. Severity of illness scoring systems. In Parrillo, J.E., Bone, R.C.(eds). *Critical Care Medicine: Principles of Diagnosis and Management*. St. Louis: Mosby, 1995.
17. Miller GA. The magical number seven, plus or minus two: Some limits in our capacity for processing information. *The Psychological Review* 1956;63:81-97.
18. Larkin J, McDermott J, Simon D, Simon H.. Expert and novice performance in solving physics problems. *Science* 1980;208:1135-42.
19. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Berger M. The APACHE III prognostic system: Risk prediction in hospital mortality for critically ill hospitalized patients. *Chest* 1991;100:1619-36.
20. Gibson RF, Haug PJ. An automated computerized severity index. *Symp Comp Appl Med Care (SCAMC)* 1994;18:332-36.
21. Oniki TA, Clemmer TP, Gardner, RM, Johnson, KV. Representative charting of vital signs in an intensive care unit. *Symp Comp Appl Med Care (SCAMC)* 1994;18:307-311.
22. Kuperman GJ, Gardner RM, Pryor TA. *HELP: A dynamic Hospital Information System*. New York: Springer-Verlag 1991.